



Nagoya City University Academic Repository

学位の種類	博士（医学）
報告番号	甲第1743号
学位記番号	第1240号
氏名	桑山 創一郎
授与年月日	令和2年3月25日
学位論文の題名	<p>Automated Detection of Macular Diseases by Optical Coherence Tomography and Artificial Intelligence Machine Learning of Optical Coherence Tomography Images (光干渉断層計（OCT）画像の機械学習を用いた黄斑疾患の自動検出に関する研究)</p> <p>Journal of Ophthalmology, 2019: 6319581, 2019</p>
	<p>View metadata, citation and similar papers at CORE.ac.uk</p> <p>田口・ 佐川 則之, 石岡 具一</p> <p>provided by NCU repository powered by CORE</p>

加齢黄斑変性 (Age-Related Macular Degeneration: AMD)、糖尿病網膜症 (Diabetic Retinopathy: DR)、緑内障は先進国における視覚障害の主要原因である。これらの診断および治療効果判定において、光干渉断層計 (Optical Coherence Tomography: OCT) は眼底の網膜構造を詳細に観察することが可能であり、眼科日常診療において必須の検査となっている。OCT は簡便で、非侵襲に、無散瞳でも測定可能で、健診のようなスクリーニングへの利用拡大に大きな期待が持たれている。従来から健診に利用されている眼底カメラと OCT を組み合わせることで、眼底疾患の検出感度の向上が期待できるが、一方で、OCT で得られた画像から正確な診断を行うためには熟練した眼科専門医の読影が欠かせない。

最近、人工知能 (Artificial Intelligence: AI) が 3 次ブームを迎え、画像・音声認識や自動運転などで実用化されてきている。ブームの背景にコンピュータの計算能力の向上に加え、機械学習の一つである畳み込みニューラルネットワークを活用した深層学習 (ディープラーニング) の登場が寄与するところが大きい。深層学習は、脳神経回路のごとく多層のレイヤーを画像の特徴を抽象化しながら下層に伝えていき、膨大な画像の入力情報と被写体名 (教師データ) を紐付け学習させることで、新規画像の特徴量から被写体を自動認識するアルゴリズムであり、医療においても画像診断への AI 導入の研究が盛んになっている。眼科領域の最初の大規模な研究として、数万枚の眼底写真を学習させ、非常に高い精度で DR を検出できたとする JAMA 誌の報告がある (JAMA 2016;316:2402-2410)。

今回、我々は、眼底カメラ以上に眼底疾患の診断に不可欠な OCT 画像の AI による自動診断が可能かを実証するため、OCT 画像をベースに深層学習を行い、黄斑疾患の自動診断の精度について検討した。

方法は、2010～2015 年に名古屋市立大学病院眼科にて撮影した OCT 画像の中から、無作為に 300 例 600 眼を抽出した。さらに各眼において、臨床で一般的に利用される中心窩 (黄斑の中心) を通る水平断および矢状断の OCT 画像を用いることとし、計 1,200 枚の画像を本研究に用いた。網膜疾患を専門とする医師が盲検的に水平断と矢状断のセットから診断を行った。この診断名を教師データとして、学習のために 1,200 枚の画像のうち無作為に 1,100 枚を抽出し、これに OCT 撮影の特徴を考慮して、回転、水平反転、平行移動を加えデータ増幅を行い、最終的に 1,100 x 54 = 594,000 枚の画像を用いて、ディープラーニングで学習させた。

次に、学習結果の評価のため、残りの 100 枚の画像を学習済みのアルゴリズムで自動診断させ、疾患候補 (正常を含む) を確信度の高い順に候補疾患を 5 つ提示させた。その中で、あらかじめ医師により診断された結果との合致率を評価した。また、各々の疾患について適合率 (Precision) と再現率 (Recall) を下記の計算式で算出し、これも評価した。

適合率 (AI 診断の正確さ) = 真陽性 / (真陽性 + 偽陽性)

再現率 (感度) (疾患検出率) = 真陽性 / (真陽性 + 偽陰性)

AI と医師の診断結果と比較した結果、100 眼中 83 眼 (83%) において、AI の候補疾患の第 1 位で医師の診断を正答していた。また、7 眼 (7%) では第 2 位の候補疾患で、4 眼 (4%) では第 3 位の候補疾患で正答が得られた。94% で候補疾患上位 3 位以内で正答していたことになる。一方、4 眼においては、5 つの候補内に医師の診断名は含まれていなかったが、嚢胞様黄斑浮腫や萎縮型 AMD のごくわずかな変化であり、診断自体が困難なものであった。

内訳の多い疾患の適合率、再現率を算出したところ、それぞれ、正常で 0.85、0.97、滲出型 AMD

で 1.00、0.77、DR で 0.78、1.00、ERM で 0.75、0.75 であった。滲出型 AMD で適合率が高く、DR では再現率（感度）が高く、疾患によって AI 検出能力に差があった。DR は網膜内の hyperreflective foci と呼ばれる特徴的所見を有していて、AI はこれにより高い感度で DR を検出していることが示唆される。一方、滲出型 AMD は特有の所見として網膜色素上皮の不規則な隆起があり、僅かな隆起所見は見逃されるものの、AI は偽陽性（誤診）なく検出することが可能のようであった。ディープラーニングは画像の特徴を客観的に独自に見出すため、これまで人間が気付きにくかった特徴を教えてくれるといった意義もありそうである。症例数を増やして検討する必要があるが、今回の結果のような精度で実装できれば、健診などでの自動スクリーニングや一般眼科医の診断支援が可能となる。健診においては、今回の「正常」の適合率 0.85 は 15% で異常を正常と診断していることとなり問題であるが、その 6 眼中 5 眼に関しては、第 2、第 3 候補で医師と一致する診断を挙げており、残りの 1 眼は非常に僅かな萎縮性変化であった。従って、上位 3 候補までを提示させることにより見逃しは防げられると思われる。

今回、我々は、比較的少ない枚数の画像をデータ増幅させ学習させることで、高い精度のアルゴリズムを構築できた。今後は希少疾患に関しても学習量を増やし、精度を向上させることで臨床応用の可能性が示された。

論文審査の結果の要旨

【背景】

加齢黄斑変性（Age-Related Macular Degeneration: AMD）、糖尿病網膜症（Diabetic Retinopathy: DR）、緑内障は先進国における視覚障害の主要原因である。これらの診断および治療効果判定において、光干渉断層計（Optical Coherence Tomography: OCT）は眼底の網膜構造を詳細に観察することが可能であり、眼科日常診療において必須の検査となっている。OCTは簡便で、非侵襲に、無散瞳でも測定可能で、健診のようなスクリーニングへの利用拡大に大きな期待が持たれている。従来から健診に利用されている眼底カメラとOCTを組み合わせることで、眼底疾患の検出感度の向上が期待できるが、一方で、OCTで得られた画像から正確な診断を行うためには熟練した眼科専門医の読影が欠かせない。

近年、人工知能（Artificial Intelligence: AI）が画像・音声認識や自動運転などで実用化されてきている。ブームの背景にコンピュータの計算能力の向上に加え、機械学習の一つである畳み込みニューラルネットワークを活用した深層学習（ディープラーニング）の登場が寄与するところが多い。深層学習は、脳神経回路のごとく多層のレイヤーを画像の特徴を抽象化しながら下層に伝えていき、膨大な画像の入力情報と被写体名（教師データ）を紐付け学習させることで、新規画像の特徴量から被写体を自動認識するアルゴリズムであり、医療においても画像診断へのAI導入の研究が盛んになっている。眼科領域の最初の大規模な研究として、数万枚の眼底写真を学習させ、非常に高い精度で糖尿病網膜症を検出できたとするJAMA誌の報告がある（JAMA 2016;316:2402-2410）。

今回、我々は、眼底カメラ以上に眼底疾患の診断に不可欠なOCT画像のAIによる自動診断が可能かを実証するため、OCT画像をベースに深層学習を行い、黄斑疾患の自動診断の精度について検討した。

【方法】

2010～2015年に名古屋市立大学病院眼科にて撮影したOCT画像の中から、無作為に300例600眼を抽出した。さらに各眼において、臨床で一般的に利用される中心窩（黄斑の中心）を通る水平断および矢状断のOCT画像を用いることとし、計1,200枚の画像を本研究に用いた。網膜疾患を専門とする医師が盲検的に水平断と矢状断のセットから診断を行った。この診断名を教師データとして、学習のために1,200枚の画像のうち無作為に1,100枚を抽出し、これにOCT撮影の特徴を考慮して、回転、水平反転、平行移動を加えてデータ増幅を行い、最終的に1,100 x 54 = 594,000枚の画像を用いて、ディープラーニングで学習させた。

次に、学習結果の評価のため、残りの100枚の画像を学習済みのアルゴリズムで自動診断させ、疾患候補（正常を含む）を確信度の高い順に候補疾患を5つ提示させた。その中で、あらかじめ医師により診断された結果との合致率を評価した。また、各々の疾患について適合率（Precision）と再現率（Recall）を下記の計算式で算出し、これも評価した。

適合率（AI診断の正確さ）＝真陽性／（真陽性＋偽陽性）

再現率（感度）（疾患検出率）＝真陽性／（真陽性＋偽陰性）

【結果】

AIと医師の診断結果と比較した結果、100眼中83眼（83%）において、AIの候補疾患の第1位で

医師の診断を正答していた。また、7 眼（7%）では第 2 位の候補疾患で、4 眼（4%）では第 3 位の候補疾患で正答が得られた。94%で候補疾患上位 3 位以内で正答していたことになる。一方、4 眼においては、5 つの候補内に医師の診断名は含まれていなかったが、嚢胞様黄斑浮腫や萎縮型 AMD のごくわずかな変化であり、診断自体が困難なものであった。

内訳の多い疾患の適合率、再現率を算出したところ、それぞれ、正常で 0.85、0.97、滲出型 AMD で 1.00、0.77、DR で 0.78、1.00、ERM で 0.75、0.75 であった。滲出型 AMD で適合率が高く、DR では再現率（感度）が高く、疾患によって AI 検出能力に差があった。DR は網膜内の hyperreflective foci と呼ばれる特徴的所見を有していて、AI はこれにより高い感度で DR を検出していることが示唆される。一方、滲出型 AMD は特有の所見として網膜色素上皮の不規則な隆起があり、僅かな隆起所見は見逃されるものの、AI は偽陽性（誤診）なく検出することが可能のようであった。

【考察】

ディープラーニングは画像の特徴を客観的に独自に見出すため、これまで人間が気付きにくかった特徴を教えてくれるといった意義もありそうである。症例数を増やして検討する必要があるが、今回の結果のような精度で実装できれば、健診などでの自動スクリーニングや一般眼科医の診断支援が可能となる。健診においては、今回の「正常」の適合率 0.85 は 15%で異常を正常と診断していることとなり問題であるが、その 6 眼中 5 眼に関しては、第 2、第 3 候補で医師と一致する診断を挙げており、残りの 1 眼は非常に僅かな萎縮性変化であった。従って、上位 3 候補までを提示させることにより見逃しは防げると思われる。

今回、我々は、比較的少ない枚数の画像をデータ増幅させ学習させることで、高い精度のアルゴリズムを構築できた。今後は希少疾患に関しても学習量を増やし、精度を向上させることで臨床応用の可能性が示された。

【審査の内容】

上記の論文要旨が申請者より発表された後、主査の芝本教授より、今回の機械学習に用いたデータ量（データ増幅の前段階）を 1,200 画像としたその根拠と妥当性、また、近年眼科領域で（プレゼンテーションに引用したもの以外の）他の機器や機械学習を用いた研究内容について、計 5 項目の質問があった。第一副査の松川教授より、眼底疾患における虚血や炎症などの病態が OCT の所見としてどのように現れてくるのか、また、学習済みの機械が最終的に行ったと思われる元画像に対しての重み付け等、計 5 項目の質問があった。第二副査の岩崎教授より、今回の教師データの診断方法（OCT 画像の読影のみによる診断結果なのか等）、また他の検査結果も加えることによる AI の診断精度の向上の可能性について、計 4 項目の質問があった。これらの質問に対して、申請者より適切な回答が得られ、学位論文の主旨を十分理解していると判断した。

本研究により、今後 AI による診断サポートシステムの構築に寄与できる可能性があり、臨床的意義も高いと考えられる。従って、本申請者は博士（医学）の学位を授与するに値すると判定された。

論文審査担当者 主査 芝本 雄太 副査 松川 則之 岩崎 真一